

Model Regresi Linear Produksi Padi di Indonesia dengan Estimasi-M

Hasih Pratiwi, Yuliana Susanti, dan Monaluvy Septiningrum

Jurusan Matematika
FMIPA UNS Surakarta
hasihprmf@yahoo.com

Abstrak

Policy of the price of main food is one of important instrument in creating resilience of national food. Remembering the importance of accomplishment effort of requirement of food, especially paddy, it is required an effort to predict production in future. There are some methods which are applicable to predict the production of paddy and to investigate the factors influencing it; one of them is regression analysis. An estimation method which is applicable to determine the regression model is M-estimation. This estimation is an extension of maximum likelihood method and robust estimation, where its estimation value is not influenced by small change in data. The purpose of this research is to determine the regression model to predict the production of paddy in Indonesia using estimation M.

Based on the result of research we conclude that the prediction model for production of paddy in Indonesia with M-estimation is

$$\hat{y} = -92,790 + 5,02x_1 + 6,37x_2.$$

The increment of one hectare farm wide and one ton seed will increase production of paddy 5.03 tons and 6.23 tons respectively. The regression parameter significance test shows that the farm wide influences the production of paddy significantly.

1. Pendahuluan

Padi adalah tanaman yang tersebar luas di seluruh dunia dan tumbuh di hampir semua bagian dunia yang memiliki cukup air dan suhu udara cukup hangat, salah satunya Indonesia. Produk dari padi, yaitu beras, menjadi makanan pokok hampir seluruh penduduk Indonesia. Oleh karena itu, stok beras harus terjaga, sinambung, dan terus ditingkatkan. Ketika stok beras dalam negeri sedang kosong, pemerintah mengambil kebijakan impor beras untuk memenuhi kebutuhan pangan penduduk. Dari sisi kepentingan konsumen beras, khususnya masyarakat kelas menengah ke bawah, kehadiran beras impor secara melimpah mendatangkan keuntungan, mengingat harganya yang relatif lebih murah daripada beras produksi dalam negeri. Namun dari sisi lain, kehadiran beras impor yang melimpah justru menjadi sebuah ancaman bagi kesinambungan produksi beras petani, terlebih saat panen raya. Sudah saatnya pemerintah mengambil kebijakan yang tidak hanya menggantungkan kebutuhan beras nasional dari mengimpor. Makalah ini membahas model regresi linear yang menggambarkan pengaruh luas lahan dan pemberian benih terhadap produksi padi di Indonesia dengan estimasi-M.

2. Model Regresi Linear

Secara umum, model regresi linear yang melibatkan p variabel bebas dapat dinyatakan sebagai

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (1)$$

dengan Y_i adalah variabel respon pada pengamatan ke- i , β_0, \dots, β_p adalah parameter, X_i adalah nilai variabel bebas pada pengamatan ke- i , dan ε_i adalah sesatan random yang berdistribusi normal, $\varepsilon_i \sim N(0, \sigma^2)$, dimana sesatannya tidak saling berkorelasi (Montgomery dan Peck, 1992: 119).

2.1. Metode Kuadrat Terkecil

Model regresi linear dapat diestimasi dengan metode kuadrat terkecil yang mempunyai ide dasar meminimumkan jumlahan kuadrat sesatan, yaitu

$$J = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2 \quad (2)$$

Untuk meminimumkan (2), dicari turunan J secara parsial terhadap β_j $j = 0, 1, \dots, p$ dan menyamakannya dengan nol sehingga diperoleh estimasi model regresi linear

$$\hat{y}_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip} \quad (3)$$

Estimasi dari sesatan disebut sisa dan dinyatakan dengan

$$e_i = \hat{y}_i - (b_0 + b_1x_{i1} + \cdots + b_px_{ip})$$

Sisa memberikan keterangan tentang data yang tidak mengikuti pola umum model regresi linear yang digunakan, ditandai dengan sisanya yang relatif besar. Sisa semacam ini disebut pencilan. Suatu data dikatakan sebagai pencilan jika $|e_i/s| > 2$ dengan s adalah rata-rata kuadrat sisa (RKS),

$$RKS = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}$$

(Sembiring, 1995: 190). Meskipun merupakan data yang tidak biasa, membuang pencilan bukanlah tindakan yang bijaksana. Bisa jadi data pencilan menyimpan informasi yang penting. Oleh karena itu, diperlukan suatu metode estimasi yang tidak peka terhadap pencilan.

2.2. Estimasi-M

Beberapa peneliti merancang suatu metode untuk mengatasi dampak dari pencilan jika digunakan metode kuadrat terkecil, yang dinamakan estimasi-M (Montgomery dan Peck, 1992: 382). Menurut Li et al. (1998), penggunaan metode kuadrat terkecil tidak akan tepat dalam menyelesaikan permasalahan yang mengandung observasi pencilan atau ekstrem (data yang tidak mengikuti pola umum data) karena asumsi kenormalan tidak dapat dipenuhi. Pada metode kuadrat terkecil, nilai sesatan yang besar akan semakin memperbesar jumlah kuadratnya. Estimasi-M mengantisipasi hal tersebut dengan mendefinisikan suatu fungsi ρ , yaitu $\rho(\varepsilon)$, yang disebut fungsi Huber. Pada estimasi-M, b_0, \dots, b_p yang secara berturut-turut merupakan estimasi dari β_0, \dots, β_p dipilih sedemikian sehingga $\sum \rho(\varepsilon)$ minimum. Fungsi Huber didefinisikan sebagai

$$\rho(\varepsilon) = \begin{cases} \varepsilon^2, & \text{jika } -k \leq \varepsilon \leq k \\ 2k|\varepsilon| - k^2, & \text{jika } \varepsilon < -k \text{ atau } \varepsilon > k \end{cases}$$

dengan $k = 1,5\hat{\sigma}$ dan $\hat{\sigma}$ adalah estimasi dari standar deviasi sisa. Untuk mengestimasi σ , digunakan $\hat{\sigma} = 1,483 MAD$, dengan MAD (median of absolute deviation) adalah median dari sisa absolut $|e_i|$ (Birkes dan Dodge, 1993: 87).

Untuk mendapatkan nilai estimasi-M, diperlukan suatu algoritma penghitungan. Berikut diberikan algoritma estimasi-M yang diacu dari Birkes dan Dodge (1993: 93).

1. Dipilih estimasi awal b_0 dari metode kuadrat terkecil lalu dihitung

$$\hat{y}_i = b_0^0 + b_1^0 x_{i1} + \cdots + b_p^0 x_{ip}$$

sisanya $e_i^0 = y_i - \hat{y}_i^0$, dan $\hat{\sigma} = 1,483MAD$.

2. Dipotong e_i^0 agar mendapatkan nilai e_i^* , yaitu:

$$e_i^* = \begin{cases} 1,5\hat{\sigma}, & \text{jika } e_i^0 > 1,5\hat{\sigma} \\ -1,5\hat{\sigma}, & \text{jika } e_i^0 < -1,5\hat{\sigma} \end{cases}$$

3. Dihitung $y_i^* = \hat{y}_i^0 + e_i^*$ lalu dari nilai y_i^* tersebut dicari nilai b_0 menggunakan metode kuadrat terkecil. Proses iterasi berlanjut sampai diperoleh nilai b_0 yang sama dengan iterasi sebelumnya.

Dari nilai b_0 di atas, diperoleh regresi dengan estimasi-M seperti Persamaan (3).

3. Uji Signifikansi Model Regresi Linear

Selanjutnya untuk mengetahui apakah variabel bebas mempunyai pengaruh yang signifikan terhadap variabel tak bebas dilakukan uji hipotesis. Hipotesisnya adalah

$$\begin{aligned} H_0 &: \beta_j = 0, \quad \forall j = q+1, q+2, \dots, p \\ H_1 &: \exists j \ni \beta_j \neq 0, \quad j = q+1, q+2, \dots, p \end{aligned}$$

dengan q adalah sejumlah variabel bebas yang masuk dalam model tereduksi

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_q X_{iq} + \varepsilon_i \quad (4)$$

Statistik ujinya adalah

$$F_M = \frac{STR_{reduced} - STR_{full}}{(p-q)\hat{\lambda}} \quad \text{dengan} \quad \hat{\lambda} = \frac{\frac{n}{m} \sum (e_i^*)^2}{n-p-1} \quad (5)$$

STR (sum of transformed residuals) adalah jumlah sisa yang ditransformasi. STR_{full} dan $STR_{reduced}$ masing-masing diperoleh dari model penuh dan model tereduksi. Algoritma STR_{full} dapat dijelaskan sebagai berikut

1. Dihitung sisa e_i dan $\hat{\sigma} = 1,483MAD$.
2. Dipotong e_i agar mendapatkan nilai $\rho(e_i)$ yaitu

$$\rho(e_i) = \begin{cases} e_i^2, & \text{untuk } -1,5\hat{\sigma} \leq e_i \leq 1,5\hat{\sigma} \\ 2k|e_i| - k^2, & \text{untuk } e_i < -1,5\hat{\sigma} \text{ atau } e_i > 1,5\hat{\sigma} \end{cases}$$

3. Dihitung $STR_{full} = \sum \rho(e_i)$.

Penghitungan $STR_{reduced}$ sama dengan STR_{full} hanya saja nilai $\hat{\sigma}$ tetap (berasal dari model penuh). Nilai $\sum e_i^*$ pada $\hat{\lambda}$ diperoleh dengan memotong e_i menjadi $-1,5\hat{\sigma}$ jika $e_i < -1,5\hat{\sigma}$ dan $1,5\hat{\sigma}$ jika $e_i > 1,5\hat{\sigma}$, sedangkan m adalah jumlah e_i yang tidak dipotong. Selanjutnya hipotesis ditolak jika $F_M > F$ tabel dengan derajat bebas $(p-q, n-p-q)$ dan tingkat signifikansi α (Birkes dan Dodge, 1993: 95).

4. Hasil dan Pembahasan

Pada penelitian ini data yang dianalisis adalah data produksi padi, luas lahan, dan banyak benih padi tahun 2006 yang diperoleh dari BPS dan Departemen Pertanian. Proses penghitungan estimasi-M yang iteratif dimulai dengan menentukan estimasi awal koefisien regresi, yang diperoleh dari metode kuadrat terkecil, yaitu

$$b^0 = (-115332; 5,02; 6,96)$$

kemudian dihitung nilai \hat{y}_i^0 dan sisa $e_i^0 = y_i - \hat{y}_i^0$. Median dari deviasi absolut $|e_i^0|$ adalah $MAD = 128227$, maka $1,5\hat{\sigma} = 285242$. Terdapat tiga sisa yang nilainya lebih besar dari $1,5\hat{\sigma}$ yaitu sisa data ke 7,9, dan 15, maka sisa ketiganya dipotong menjadi 285242. Lima sisa yang lebih kecil dari $-1,5\hat{\sigma}$, yaitu sisa data ke 2,6,8,20, dan 22, dipotong menjadi -285242 dan diperoleh nilai $y_i^* = \hat{y}_i^0 + e_i^*$. Nilai y_i^* yang diperoleh digunakan untuk menghitung b^0 pada iterasi kedua.

Pada iterasi kedua, diperoleh $MAD = 145840$ sehingga $1,5\hat{\sigma} = 324421$. Tidak ada nilai sisa yang dipotong, sehingga nilai y tidak berubah. Hal ini mengakibatkan estimasi model regresi sama dengan sebelumnya dan iterasi berhenti. Jadi, model regresi linear dengan estimasi-M adalah

$$\hat{y} = -92790 + 5,02x_1 + 6,37x_2 \quad (6)$$

Model regresi (6) menunjukkan bahwa untuk peningkatan setiap satu hektar luas lahan padi, maka produksi padi juga meningkat sebesar 5,02 ton. Untuk banyak benih yang meningkat satu ton akan mengakibatkan peningkatan produksi padi sebesar 6,37 ton. Setelah diperoleh model regresi linear (6), dilakukan uji hipotesis untuk mengetahui apakah luas lahan dan banyak benih mempunyai pengaruh terhadap produksi padi di Indonesia tahun 2006. Ada tiga kemungkinan model tereduksi pada model regresi linear dengan dua variabel bebas, yaitu

$$Y_i = \beta_0 + \varepsilon_i \quad (7)$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \quad (8)$$

$$Y_i = \beta_0 + \beta_2 X_{i2} + \varepsilon_i \quad (9)$$

Tabel 1: Estimasi Model-Model Tereduksi tiap Iterasi

Iterasi ke	b^0		
	$Y_i = \beta_0 + \varepsilon_i$	$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$	$Y_i = \beta_0 + \beta_2 X_{i2} + \varepsilon_i$
1	(1075340)	(-121910;5,15)	(337364;203)
2	(953402)	(-101994 ; 5,16)	(247212 ; 204)
3	(923841)	(-99950 ; 5,16)	(226408 ; 204)
4	(916505)	(-99641 ; 5,16)	(220989 ; 204)
5	(914283)	(-99594 ; 5,16)	(219512 ; 204)
6	(913609)	(-99587 ; 5,16)	(219095 ; 204)
7	(913405)	(-99585 ; 5,16)	(218974 ; 204)
8	(913343)	(-99585 ; 5,16)	(218939 ; 204)
9	(913324)		(218928 ; 204)
10	(913319)		(218925 ; 204)
11	(913317)		(218924 ;204)
12	(913316)		(218923 ; 204)
13	(913316)		(218923 ; 204)

Sebelum dilakukan uji hipotesis untuk masing-masing model tereduksi, dihitung nilai STR_{full} yang berasal dari model regresi (6). Berdasarkan algoritma penghitungan STR_{full} , langkah pertama yang dilakukan adalah mencari sisa e_i dari model regresi (6). Diperoleh nilai $MAD = 145840$ sehingga $k = 1, 5\hat{\sigma} = 324421$. Semua sisa berada pada interval $-1, 5\hat{\sigma} \leq e_i \leq 1, 5\hat{\sigma}$ dan diperoleh nilai $\rho(e_i) = e_i^2$. Hasil penjumlahan $\rho(e_i)$ adalah $STR_{full} = (1, 044562)10^{12}$. Untuk mendapatkan nilai $\hat{\lambda}$, sebelumnya dihitung nilai $\sum(e_i^*)^2 = (1, 044562)10^{12}$ dan diperoleh

$$\hat{\lambda} = \frac{\frac{n}{m} \sum(e_i^*)^2}{n - p - 1} = \frac{\frac{33}{33}(1, 044562)10^{12}}{33 - 2 - 1} = 34818733333$$

Selanjutnya dihitung $STR_{reduced}$ yang berasal dari model tereduksi (7), (8), dan (9). Namun terlebih dahulu dicari estimasi ketiga model tereduksi, sama seperti model penuh tetapi nilai $\hat{\sigma}$ tetap (tidak iteratif, berasal dari model penuh), sehingga $1, 5\hat{\sigma} = 324421$. Estimasi model tereduksi (7), (8), dan (9) disajikan dalam Tabel 1.

Kemudian dilakukan uji hipotesis untuk masing-masing model tereduksi. Hipotesis untuk model tereduksi (7) adalah

$$H_0 : \beta_i = 0, \forall i = 1, 2$$

luas lahan dan banyak benih tidak mempunyai pengaruh yang signifikan terhadap produksi padi)

$$H_1 : \exists i \ni \beta_i \neq 0 i = 1, 2$$

(salah satu dari luas lahan atau banyak benih mempunyai pengaruh yang signifikan terhadap produksi padi)

Statistik ujinya adalah

$$F_M = \frac{STR_{reduced} - STR_{full}}{(p - q)\hat{\lambda}} = \frac{(1,687020)10^{12} - (1,044562)10^{12}}{(2 - 0)(34818733333)} = 9,22576$$

Nilai F tabel dengan tingkat signifikansi 0,05 dan derajat bebas $(p - q, n - p - q) = (1, 30)$ adalah 4,17. Berdasarkan uji signifikansi koefisien regresi, dapat disimpulkan bahwa H_0 ditolak karena $F_M = 9,22576 > F$ tabel. Jadi, paling tidak satu dari luas lahan atau banyak benih mempunyai pengaruh yang signifikan terhadap produksi padi.

Untuk memastikan variabel apa saja yang mempunyai pengaruh yang signifikan terhadap produksi padi, dilakukan uji untuk model tereduksi yang lain. Hipotesis untuk model tereduksi (8) adalah

$$H_0 : \beta_2 = 0$$

(banyak benih tidak mempunyai pengaruh yang signifikan terhadap produksi padi)

$$H_1 : \beta_2 \neq 0$$

(banyak benih mempunyai pengaruh yang signifikan terhadap produksi padi)

Statistik ujinya adalah

$$F_M = \frac{STR_{reduced} - STR_{full}}{(p - q)\hat{\lambda}} = \frac{(1,121191)10^{12} - (1,044562)10^{12}}{(2 - 1)(34818733333)} = 2,20079$$

Nilai F tabel dengan tingkat signifikansi 0,05 dan derajat bebas $(2, 30)$ adalah 3,32, sehingga dapat disimpulkan bahwa H_0 tidak ditolak karena $F_M < F$ tabel. Jadi, banyak benih tidak mempunyai pengaruh yang signifikan terhadap produksi padi.

Hipotesis untuk model tereduksi (9) adalah

$$H_0 : \beta_1 = 0$$

(luas lahan tidak mempunyai pengaruh yang signifikan terhadap produksi padi)

$$H_1 : \beta_1 \neq 0$$

(luas lahan mempunyai pengaruh yang signifikan terhadap produksi padi)

Statistik ujinya adalah

$$F_M = \frac{STR_{reduced} - STR_{full}}{(p - q)\hat{\lambda}} = \frac{(1,811826)10^{12} - (1,044562)10^{12}}{(2 - 1)(3481873333)} = 22,036$$

Jadi, H_0 ditolak karena $F_M = 22,036 > F$ tabel. Artinya, luas lahan mempunyai pengaruh yang signifikan terhadap produksi padi. Ketiga hipotesis tersebut menghasilkan suatu kesimpulan yaitu bahwa hanya luas lahan yang memberikan pengaruh signifikan terhadap produksi padi.

5. Kesimpulan

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa model regresi produksi padi di Indonesia dengan estimasi-M adalah

$$\hat{y} = -92.790 + 5,02x_1 + 6,37x_2$$

Peningkatan setiap satu hektar luas lahan padi dan satu ton benih, masing-masing akan meningkatkan produksi padi sebesar 5,02 ton dan 6,37 ton. Berdasarkan uji signifikansi parameter regresi disimpulkan bahwa luas lahan mempunyai pengaruh yang signifikan terhadap produksi padi.

Pustaka

- [1] Badan Pusat Statistik (2006). Harvested Area, Yield Rate and Production of Paddy by Province. <http://www.bps.go.id>.
- [2] Birkes, D. and Y. Dodge (1993). Alternative Method of Regression. John Wiley and Sons, Inc. New York.
- [3] Departemen Pertanian (2006). Benih Padi. <http://www.deptan.go.id>.
- [4] Li, S.Z., Wang, H., and Soh, W.Y.C. (1998). Robust Estimation of Rotation Angle from Image Sequences Using the Annelling M Estimator, Jurnal of Mathematical Imaging and Vision, Vol. 8. No. 2, 181-192.
- [5] Montgomery, D. C. and E. A. Peck (1992). Introduction to Linear Regression Analysis, 2nd ed. John Wiley and Sons, Inc. New York.
- [6] Sembiring, R. K. (1995). Analisis Regresi. Penerbit ITB. Bandung.